DOT/FAA/AM-99/8

Office of Aviation Medicine
Washington, D.C. 20591

Optimizing Blink Parameters For Highlighting an Air Traffic Control Situation Display

Nelda J. Milburn
Henry W. Mertens
Civil Aeromedical Institute
Federal Aviation Administration
Oklahoma City, Oklahoma 73125

March 1999

Final Report

U.S. Department
of Transportation

Federal Aviation Administration

NOTICE

Technical Report Documentation Page

1. Report No. DOT/FAA/AM-99/8	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Optimizing Blink Parameters for Highlighting an Air Traffic Control Situation Display		5. Report Date March 1999
		6. Performing Organization Code
7. Author(s) Milburn, N.J., and Mertens, H.W.		8. Performing Organization Report No.
9. Performing Organization Name and Address FAA Civil Aeromedical Institute P.O. Box 25082 Oklahoma City, OK 73125		10. Work Unit No. (TRAIS)
		11. Contract or Grant No.
12. Sponsoring Agency name and Address Office of Aviation Medicine Federal Aviation Administration 800 Independence Avenue, S.W. Washington, D.C. 20591		13. Type of Report and Period Covered
		14. Sponsoring Agency Code

15. Supplemental Notes

This work was performed under task HRR-190.

16. Abstract

Research suggests blinking targets are more alerting than steady targets; however, several factors can interact with blinking to either improve or degrade its attention-getting value. Those factors include target size, color, brightness contrast, frequency of blink, and the time the blink is at maximum brightness relative to the time it is off or dim during the blink cycle. No guidelines were found for determining the optimal attention- getting blink amplitude (the percentage of decrease in target brightness from a standard) and the interaction of blink amplitude with the other blink characteristics mentioned above. Thirty-six participants were asked to locate and select blinking blocks of text on a simulated air traffic control display to examine the interaction of blink frequencies, amplitudes, and durations with size of text. Our results support the use of amplitudes 75% or greater combined with frequencies from 2 to 4 Hertz and text size 0.15 inch or greater for optimum highlighting value in visual search tasks.

17. Key Words Attention-Getting, Text Size, Target, Flashing CRT Display, Frequency, Air Traffic Control, Amplitude, Blink	18. Distribution Statement Document is available to the public through the National Technical Information Service, Springfield, Virginia 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	26	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

TABLE OF CONTENTS

Optimizing Blink Parameters for Highlighting An Air Traffic Control Situation Display

INTRODUCTION

A primary consideration of any use of computers is the presentation of information flowing *from* the computer *to* the human in a concise form that is quickly and easily interpreted. The varied research that is inspired by that topic sentence is extensive, including screen design, computer messages, computer response time, types of output, types of displays, resolution, refresh rates, luminance and contrast just to name a few. Specifically, this study examined the effectiveness of blinking or flashing of text as a method of gaining the user's attention.

Though researchers and display designers cannot agree on a term for the design feature (e.g. blinking, flashing, flickering, winking, pulsing, and off-on cycling), several sources (Anastasi, Hill, Murphy, Cardosi, Guttman, & Amaldi, 1995; Boff & Lincoln, 1988; Christ, 1975; Gerathewohl, 1951, 1952, 1953, 1954; Gilmore, Gertman, & Blackman, 1989; Military Standards, 1989; Thackray & Touchtone, 1991; Van Orden & Di Vita, 1993) agree that blinking targets are more alerting than steady-state targets and can aid the user in finding the targets quickly. Those sources have recommended information coding dimensions such as target size, color, shape, brightness contrast, and frequency, but no guidelines have been recommended for a minimum blink amplitude that is most effective for attention-getting. Because those studies involved blinking characteristics in which the targets alternated between on and off (100% change in intensity) a recommendation of 100% amplitude change during blinking was implicit.

A previous study (Mertens & Milburn, 1998) evaluated the effectiveness of redundant color coding for protecting the performance of individuals with color vision deficiency in a simulated air traffic control search task requiring search and identification of specific aircraft. Blinking was used along with color-coding as one of the redundant cues. The blinking characteristics of 20% amplitude modulation, 2 hertz (HZ) blink rate, and .1 sec blink duration were used in that study to simulate those used in the developmental Initial Sector Suite System (ISSS), a system now discontinued, that was to provide a new work station for controllers. Other coding dimensions and conditions of that study also simulated conditions of the ISSS. We do not know why those characteristics of blinking were selected for the prototype displays. (Those characteristics of blinking may have been selected to reduce the potential distraction effects of blinking, or to reduce a possible adverse effect of blinking on legibility.) The results of our study demonstrated that the 20% decrease in brightness intensity at a 2 HZ rate was of little or no attentional value for both individuals with normal and abnormal color vision (Mertens & Milburn, 1998).

As mentioned above, our initial interest in blinking was based on examination of early prototypes of new air traffic control displays and the use of blinking as a redundant cue used with color. Our primary research concern deals with development of guidelines for use of color on CRT displays that would allow color deficient users to adequately use the color-coded information. Guidelines for design of electronic displays recommend that color coding should always be used as a redundant cue, such that some other characteristic of the symbol, shape, size, brightness, alphanumerics, blinking, and so forth, conveys the same information (Silverstein & Merrifield, 1985). One of the stated benefits of redundant color-coding is that it "...permits people with color vision deficiencies (CVD) to interpret color coded displays" (Society of Automotive Engineers, 1988). Redundant color-coding can have functions other than coding, such as decluttering information and attention-getting. When the use of color is for the latter purposes, the attention-getting value of the redundant cue or cues must be considered. Therefore, we conducted a study that focused on the attention-getting value of blinking amplitudes to determine how much of a change in brightness was necessary to reliably capture the user's attention if no other redundant cues were available. Our intent was

to use the resulting information in the preparation of another study examining the use of blinking as a redundant cue with color.

Our first study on blink amplitudes (Milburn & Mertens, 1997) used an air traffic control display similar to the one used in this experiment except that only the amplitude variable was manipulated. For that study, target blink amplitude was defined as the percentage of decrease in luminance from a standard of 51.4 Cd/m^2 at a rate of 2 HZ, and the duration of the blink intensity decrement was approximately 1/10 sec. The 7 amplitudes of blink were 100%, 93.75%, 87.5%, 75%, 50%, 25%, and 12.5%. Some blink amplitudes had such poor attention-getting value that 2 respondents were unable to find even a single target of the 26 possible during an 18-trial experimental condition. On the average, participants made very few false alarms for any of the different blink amplitude levels. Most errors were misses, defined as, a failure to detect and select a blinking target. When the targets were difficult to detect, some respondents terminated the trial without finding any targets, or without finding all of the targets on the trials that contained multiple targets. High miss rates and low false alarm rates were typical in conditions involving a small change in amplitude ranging from 12.5% to 25%. Of the types of errors discussed, misses are the most critical in air traffic control work. A false alarm might result in the air traffic controller making some additional mental calculations, with a loss of time, but failure to quickly find an emergency warning or critical event displayed on the ATC situation display could result in a loss of required separation between aircraft.

Performance in that study was stable with high accuracy ranging from 98.75% to 100% correct for conditions involving a 75% to 100% decrease in brightness but showed a steady increase in errors (missed targets and selection of non-targets) for the 50%, 25%, and 12.5% decreases in brightness. Performance in conditions involving amplitudes of 25% or less ranged from 5% correct to 72.5% correct.

The objective of Milburn and Mertens' (1997) study was to determine the level of blink amplitude that can capture the user's attention quickly and with a high degree of regularity (at least 95% of the time). Note that others have arbitrarily used 95% correct criterion for selection of cue parameters in the design of symbology for aviation systems (Silverstein & Merrifield, 1985). Milburn and Mertens (1997) found a strong relationship between accuracy and response time under each of the blink amplitude conditions. Unwavering accuracy and response time performance were maintained for amplitudes ranging from 75% to 100%, and a steep decline in accuracy, combined with increased response times, was evident for the amplitudes ranging from 12.5% to 25%.

Results from Milburn and Mertens (1997) aided the design of this study and our attempt to address the complex issues surrounding the selection of optimal attention-getting blink parameters. The studies are alike in that similar methodologies were used; however, this experiment addressed a more complex set of parameters relevant to designing ATC situational displays. The first experiment examined a range of blink amplitudes in order to determine whether performance differences exist between blink amplitudes. We selected 4 blink amplitudes (25%, 50%, 75% & 100%) that were used in the first study to use in a more complex design intended to examine the interaction of blink amplitude with other blink characteristics such as frequency, duration and text size. This study expanded the design of the first by adding frequency as a factor (1, 2, 3, & 4 Hertz), comparing 2 blink durations (off or dim times of 0.1 sec and one-half of the blink cycle), and comparing 3 sizes of text (character heights of 0.10, 0.15, & 0.20 inch).

Concisely stated, the purpose of this study was to examine the interaction of blink amplitude with frequency, duration and size of text and to select optimal attention-getting blink parameters from among those tested.

METHOD

Participants

Prior approval for all procedures and use of human participants was obtained from the Office of Aviation Medicine/Civil Aeromedical Institute human use committee. Volunteers were recruited by the Human Resources Research Division of the Civil Aeromedical Institute. The informed consent of every participant was obtained prior to participation, and each participant was free to withdraw from the experiment without prejudice at any time during the experiment.

Thirty-six participants ranging in age from 18 to 34 years of age served as participants in this study. Participants were matched by gender and age for each of the between-subject groups (text size and duration). Eighteen of the participants were male. Because

of known response time differences related to age, participants were matched by age for the six groups. An ANOVA was conducted to determine the effectiveness of the placement of participants to design cells with regard to age. No significant differences were found. Table 1 presents participant ages and mean age for each group. Participants placed in the small and medium text groups had a mean age of 23.3 years and the large text group had a mean age of 25.5 years. The people who served as subjects in the .1 sec duration group had a mean age of 23.9 years and those in the half-cycle group averaged 24.2 years old.

All volunteers had at least 20/30 corrected visual acuity in both near and distant vision as determined with the Bausch and Lomb Orthorater. Normal color vision was determined by the Ishihara's Test for Color-Blindness 24-Plates Edition, and Farnsworth's F-2 PIC Plate test.

Design

The mixed model experimental design included 4 blink frequencies (1, 2, 3, & 4 Hertz), 4 Blink amplitudes (25%, 50%, 75%, & 100% reduction in brightness), 2 blink durations (off or dim times of 0.1 sec and one-half of the blink cycle), and 3 sizes of text (character heights of 0.10, 0.15, & 0.20 inch). Amplitude and frequency were repeated measures, and duration and size were between-subjects factors. Table 2 provides a graphical representation of the design and number of participants within each cell.

Procedure

The ATC situation display was presented on a Tektronics 4125 19-inch monitor and consisted primarily of aircraft symbols, each with a block of alphanumeric data attached by a leader line to the aircraft symbol. The alphanumeric data, called the data block, contained information about the state of the aircraft, e.g., call sign, aircraft type, heading, altitude, and ground speed. The aircraft symbols and their data blocks were positioned throughout the situation display. Map data that designated boundaries of air traffic control sectors, geographic data, and navigation references were also included in the situation display. Participants were instructed to locate all of the blinking data block targets, to select them using a mouse, and to press the "enter" key to advance to the next trial. They searched for 1 or 2 targets, within a field of 16 aircraft (data blocks), on each of the simulated ATC situation display. Location of targets was randomized in each of 8 areas or divisions of the screen with each combination of areas appearing once in 2 target trials. Data blocks other than the blinking targets were called distracters.

Each experimental condition consisted of 1 demonstration trial to illustrate the amplitude of the blink for the targets in that condition, followed by 5 practice trials and 18 test trials. Fourteen trials contained 2 targets, and 4 trials contained 1 target. Participants were informed that some trials would present screens with only 1 target. Targets were not redundantly coded, so identification of targets was based on the participant's perception of the blinking parameters.

A control condition was presented to obtain a baseline and an estimation of the optimum response time. The control condition consisted of trials with display screens that were identical to those presented in the experimental conditions, except that they contained no distracter targets. Each screen contained either 1 or 2 data blocks. All data blocks on the screen were steady, but were considered the targets, and the participant's task was to select them as quickly as possible. The targets were displayed on the same background map as used for trials with distracters, and the locations of targets were varied by the same techniques as for trials with distracters.

The 4 amplitudes of blink were 100%, 75%, 50%, and 25%. For this study, target blink amplitude is defined as the percentage of decrease in luminance from a standard luminance of 51.4 Cd/m^2. Frequencies of 1, 2, 3, and 4 Hertz (HZ) were used. The duration of the blink intensity decrement was either 1/10 sec or half of the blink cycle time (half-cycle).

The instructions urged the participants to respond quickly and accurately, with the emphasis on accuracy. All conditions were self-paced and four 5-minute breaks were programmed into each participant's unique random order of conditions, so that the breaks were evenly spaced throughout the experiment. All participants completed the experiment within 4 hours.

Performance Measures

Performance was measured in terms of accuracy (percent correct) and search time. Search errors included misses (failure to select a target), and false alarms (selection of non-targets). The response time measure was defined as the elapsed time between the onset of the trial and the subject's key-press-termination of the trial.

RESULTS

The SPSS General Linear Model (GLM) repeated measures procedure was used to analyze the data. All multivariate tests reported used Pillai's criterion because that test pools the statistics from each dimension to test the effect, and as a result, is more robust to violations of the assumption of homogeneity of variance-covariance matrices. The data were analyzed separately for search errors (percent correct) and response time, and then summarized considering the combination of performance measures and blink parameters relative to optimal performance observed under the control condition.

The primary purpose of these analyses was not to detect an effect due to one of the repeated measures or the between-groups factors; but rather to isolate the conditions among which there are no differences in performance. As a result of isolating those conditions, we can identify the combination of blink parameters that will consistently capture the attention of the user to promote optimal performance.

Control Condition Analysis

One participant made 1 error in the control condition. Our supposition concerning that error is that it was an accident or a lapse in attention. The control condition did not involve searching for targets among distracters, so it is hypothesized that it was not a search or identification error. The response times for the control condition were very short so a number of repetitions occurred in a small amount of time making it likely that the individual did not verify that the target was selected before terminating the task. Assuming that it was an accident, this information gives us some idea of the probability of such accidental errors happening in the experimental conditions. That one error occurred from a total of 648 responses (18 trials times 36 participants), or less than .002 percent of the time. The mean percent correct for the 36 participants was 99.802 with a standard deviation of 1.19. No significant response time differences between the six groups in the control condition were found using an ANOVA. Table 3 includes the mean response time for each group under the control condition, which ranged from 3.98 sec to 5.23 sec. with an overall average of 4.75 sec.

Analysis of Search Errors

Frequency and amplitude. Both frequency [$F_{3,28}$ =7.65,a = .05, p= .001] and amplitude [$F_{3,28}$ =8.16,a =. 05, p< .001] were significant as main effects in error data analysis, and the interaction between frequency and amplitude was significant [$F_{9,22}$ =3.04,a= .05, p= .016]. Figure 1 illustrates the lower percent correct responses under the 1 Hertz condition and the gradually improving percent correct for 2, 3, and 4 Hertz conditions. The 25% amplitude condition showed the most improvement (increase of percent correct) with the faster frequencies. However, the poor attention-getting attributes of the 25% and 50% amplitudes are evident across frequencies when compared to the larger amplitude conditions.

Size and Duration. Table 4 presents the percent correct averaged over the 2 blink durations because duration was not found to be a significant factor in the analysis involving the full model (4 levels of frequency, 4 levels of amplitude, 3 sizes, and 2 durations). The percent correct ranged from 22.22% for the smaller amplitudes and slower frequencies with small and medium text to an average of 100% correct in the most attention-getting conditions. In all frequency conditions, the 25% amplitude conditions had a lower percent correct compared to the larger amplitudes for that frequency. It is important to note that in each experimental condition at least 1 person was able to achieve a 100% correct score; therefore, inclusion of the maximum range column on Table 4 was unnecessary.

Table 4 was included primarily to provide information relevant to the distribution of accuracy scores under each condition. For example, in the most attention-getting conditions, the standard deviation for participants ranged between 0 and 2.16. However, the standard deviation among participants was as high as 30.0 for the 2 HZ, 25% amplitude condition with small text. The minimum scores have been included for each condition to demonstrate the range of performance and to facilitate selection of blink parameters that will accommodate most users. Cells in which the minimum performance is less than 90% correct have been shaded on Table 4 to serve as a visual aid relevant to the task of establishing a cut-off point for purposes of making recommendations for blink parameters. Previous research (Milburn &

Mertens, 1997; Mertens & Milburn, 1998) indicated that a blink amplitude of 25% is not effective as an attention-getting mechanism; and, conversely, ample research (Anastasi, Hill, Murphy, Cardosi, Guttman, & Amaldi, 1995; Boff & Lincoln, 1988; Christ, 1975; Gerathewohl, 1951, 1952, 1953, 1954; Gilmore, Gertman, & Blackman, 1989; Military Standards, 1989; Thackray & Touchtone, 1991; Van Orden & Di Vita, 1993) has demonstrated that an on/off blink (100% amplitude) is effective. In a study (Milburn & Mertens, 1997) with blink parameters of 2 HZ, .1 sec blink duration, and 7 blink amplitudes ranging from 12.5% to 100%, performance was stable for the 75% amplitude condition and greater, and was clearly unacceptable for conditions of blink amplitude less than 50%. In that study, a paired-sample t-test was used to determine whether a significant difference existed between the 50% and 75% amplitude conditions. No significant difference was found, but a conservative recommendation to avoid amplitudes less than 75% was made based on the small sample size of that study and the marginally acceptable performance obtained for the 50% amplitude condition. Therefore, based on results of that study and because clear-cut dividing lines are not readily apparent in this study's data, further analyses were conducted to provide more information pertinent to making blink parameter recommendations.

Model Comparisons. The omnibus test described above was conducted to determine if performance differed as a result of varying frequency, amplitude, duration and text size, and many differences were found. The omnibus test answers the question, "Are any (performance) differences present among the levels of the independent variables?" However, it cannot identify which levels are significantly different from another level, and under which levels the performance is the same based on the resulting F-statistic. Therefore, in an effort to isolate the conditions with essentially the same performance (near optimal) from sub-optimal conditions, a test of focused contrast was used. That technique involves analyzing a sub-set of a complex design by systematically omitting levels of a factor from the analysis. This is particularly useful when the number of levels is large or the design is complex. The levels represent a range of conditions (1 to 4 Hertz, and 25% to 100% amplitude) so the strategy was to eliminate the minimum level first, analyze the remaining factors to test for a difference, and if found, then proceed by dropping the next higher level. The final step was to test adjacent levels of a factor to find break points at which performance differs. At least with this experiment, it would be advantageous from a CRT display design standpoint if a "cut-point" existed above which performance was stable and near optimal and below which performance was clearly unacceptable. Realistically, that cut-point is usually not well defined, especially in complex designs and if more than one dependent measure is involved. The description of the analyses that follow was an attempt to discover that cut-point for each of the two dependent variables.

As mentioned above, the first step was to conduct an omnibus test and to examine the results. The descriptive statistics of this study were consistent with the results from the pilot research. Evidence from those two studies indicates that the 25% amplitude condition produced unacceptable performance relative to the other amplitudes. See Figure 2. Consequently, that level of amplitude was dropped from further analysis. Subsequent analyses of the response accuracy measure focused on evaluating the 3 remaining levels of amplitude and 4 levels of frequency.

The second step of the model-comparison procedure involved an analysis of the 4 frequencies, 3 text sizes and the 50% and 75% amplitude conditions, to answer the hypothesis question related to the effectiveness of the 50% amplitude condition. Significance was found for frequency [$F_{3,28}$ =5.05,a= .05, p= .006], amplitude [$F_{1,30}$ =4.1,a= .05, p= .052], frequency by amplitude [$F_{3,28}$ =4.64,a=. 05, p= .009], and a between-groups effect due to size [$F_{2,30}$ =4.89,a= .05, p= .015]. Figure 2 illustrates the interaction between frequency and amplitude; likewise, Figure 3 graphs the percent correct measure as a function of frequency, amplitude, and text size.

Step 3 was conducted to determine if a significant difference existed between the 50% and 75% amplitude conditions if the 1 HZ condition were dropped. Therefore, another model comparison procedure was run that involved only those 2 amplitudes and 3 frequencies (2, 3, and 4 HZ) and 3 text sizes. The results revealed a significant interaction between frequency and amplitude [$F_{2,29}$ =4.2,a= .05, p= .025], no significant between-groups factors, and no significant post-hoc multiple comparisons for the size factor.

Finally, in an analysis involving 2, 3, and 4 HZ, 3 text sizes and the 75% and 100% amplitude conditions; amplitude, frequency, and text size were no longer significant as main effects or as interaction effects. The results of Step 4 suggest that the larger amplitudes of a blinking target have potent, attention-getting value in a search task. However, an accuracy performance difference related to the duration of the blink emerged. Table 5 shows no errors were made under the .1 sec duration condition at 2, 3, and 4 HZ with 75% and 100% amplitudes, and a total of 10 errors was made by the 18 people who served as participants under the half-cycle blink duration condition. Most errors occurred under the 2 and 3 HZ conditions; only 1 participant made 1 error on the 18 trials presented under the 4 HZ frequency.

A summary (Table 6) of these model comparison results is provided to aid the reader in understanding the blink parameters that produced essentially the same accuracy performance and those that did not.

Analysis of Search Response Times

Initially, the full model was analyzed, and the following 4 sections briefly discuss the findings relevant to main effects and interaction effects of frequency, size, duration and amplitude. Following those individual topics, a discussion of the model comparisons procedure similar to the one conducted on the accuracy data is addressed for the search response time analysis.

Frequency. Frequency of blink was significant [$F_{3,28}=49.23$,a=. 05, $p< .001$] and is especially noticeable in the pattern of response times for the 25% amplitude across different frequency conditions (Figure 4). As the frequency increased, the response times gradually decreased for all amplitudes, with the 25% amplitude conditions showing the most reduction in response time at the 4 HZ frequency.

Size. Size of text was significant as a between-groups factor [$F_{2,30}=3.66$,a= .05, $p= .038$] and there were significant interactions of size with frequency [$F_{6,58}=3.51$,a= .05, $p= .005$], and size with amplitude [$F_{6,58}=5.02$,a= .05, $p< .001$]. These interactions can be seen in Figure 5. The slower response times at the 25% amplitude, compared to the 75% and 100% amplitudes, combined with the distinct pattern (slow to faster) response times for small, medium, and large text across frequencies is evident in Figure 5. The consistent pattern of responses found in this study supports the theory that size of target text is important if the blink frequency is infrequent (does not occur during a typical dwell time during a visual search). Furthermore, if the amount of change in brightness is subtle, as noted by the longer search times in the 1HZ-25% amplitude condition, the large text has a definite advantage over the smaller sizes.

Duration. In the analysis involving the full-factorial model (4 frequencies, 4 amplitudes, 2 durations, and 3 text sizes), duration was not significant. See Figure 6.

Amplitude. As a main effect, amplitude was significant [$F_{3,28}=22.79$,a= .05, $p< .001$] with the greatest mean response-time differences under the 25% amplitude condition. Figures 4 and 5 portray the disparity between the amplitude conditions. There are very small mean response time differences between the 75% and 100% (.05 sec) amplitude conditions, with larger differences between the 50% and 75% amplitudes (.79 sec), and the greatest difference is between 25% and 50% amplitudes (4.68 sec). It is clearly evident that for most applications, a .05 sec response time difference (between 75% and 100% amplitudes) has little or no significance from a practical application viewpoint. Furthermore, using that same criterion to determine practical significance, a 4.68 sec difference between 25% and 50% amplitudes may be consequential in time-critical, safety applications. However, when that same question is applied to the 50% and 75% interval, the interpretation of significance is more difficult. Statistically, the .79 sec difference is significant, [$F_{1,30}=20.92$,a= .05, $p< .001$] but depending upon the application to which these results are generalized, acceptability will obviously depend upon the criticality of that length of mean response time and the response accuracy associated with the amplitude condition.

Descriptive Statistics for Response Time.

Table 7 presents the descriptive statistics (minimum, maximum, mean and the standard deviation) for response time under the experimental and control conditions. There are many ways of evaluating the response time performance of the experimental conditions. Bailey (1989) recommends designs that will accommodate 95% of the users by considering the full range of measurements. If that advice is applied to the response time data and the control condition is used as a baseline of response time, the 95th percentile response time is 8.16 sec. It seems reasonable that the mean for the experimental conditions should be

as good as the 95th percentile response time for the control condition. Therefore, the cells with means greater than 8.16 sec have been shaded on Table 7. Granted, this may be a very liberal initial criterion and may not represent ideal performance, but it will serve as a starting point for eliminating conditions with the longest response times. Also, many of the same experimental conditions are shaded on both Tables 4 and 7.

Model Comparisons. A 4-step model comparison procedure similar to the one described for the accuracy data was used on the response time data. Once again, the purpose of these analyses was to uncover a "cut-off point" at which the performance is statistically significantly different from the remaining levels of a factor, and to identify levels of a factor that exhibit essentially identical and optimal performance. The technique used is analogous to the 4 steps performed on the search accuracy data. The full model analysis results described above constituted Step 1.

The analysis of Step 2 involved 4 frequency conditions, the 50% and 75% amplitude conditions and size as a between-groups variable. Significant effects were found for frequency [$F_{3,28}$ = 10.9,a= .05, $p<$ .001], amplitude [$F_{1,30}$ = 29.09,a= .05, $p<$ .001], amplitude by size [$F_{2,30}$ = 12.84,a=.05, $p<$ .001], and frequency by amplitude [$F_{3,28}$ = 3.45,a=.05, $p=$ .03].

Response times for the 4 frequency conditions and the 50% and 75% amplitude conditions averaged over duration were compared using a Games and Howell multiple comparison procedure to control alpha for 3 comparisons (at each frequency) to determine if significant differences exist between the 3 text sizes. Response time differences were significant comparing the large with the small and medium text for the 50% amplitude conditions both at the 1 HZ and 3 HZ frequencies and the large with medium text for 1 HZ and 75% amplitude.

Step 3 of the analysis involved dropping the 1 HZ level of the frequency factor and making the first comparison of adjacent levels of the amplitude factor. The size and duration factors remained in the model. Significant factors, interactions and post-hoc comparisons were undifferentiated from the significant results of Step 2 with the following exceptions. First, because the 1HZ level was deleted from this analysis, the post-hoc multiple comparisons do not appear in the list of significant comparisons. Second, the interaction of frequency by amplitude was not present in Step 3 of the analysis. However, frequency [$F_{2,29}$ = 5.75,a= .05, $p=$ .008], amplitude [$F_{1,30}$ = 20.92,a= .05, $p<$ .001], and an interaction of amplitude with size [$F_{2,30}$ = 11.06,a= .05, $p<$ .001] remained on the significance list.

Finally in Step 4, response times for the 2, 3, and 4 HZ, 75% and 100% amplitudes, 2 durations, and 3 text sizes were analyzed. The only significant effect was due to frequency [$F_{2,29}$ = 12.36,a= .05, $p<$ .001]. That significance is probably a result of the slightly faster response times under the 4 HZ conditions. The mean response times for 2, 3, and 4 HZ are 5.76, 6.05, and 5.36 sec respectively. Once again, small differences in mean response time (.4 and .69 sec) for the 4 HZ condition as compared to the 2 and 3 HZ are statistically significant; but practical implications should be considered. Furthermore, there were no significant Games and Howell post-hoc comparisons for the text size variable.

Table 8 provides a concise summary of the significant results found in the 4-step model-comparisons procedure for the response time dependent measure.

Summary of Results

Blinking has high attention-getting value at 2, 3, and 4 Hertz when the decrease in brightness is 75% or greater; accuracy and speed of search were reduced at lower amplitudes. The recommended duration of a blink (off or dim time) is typically half of the cycle time; however, our research indicates that for frequencies of 2 to 4 Hertz, a shorter 0.1 sec blink was as effective as blink durations that are half of the blink period. Size of text was statistically significant as a between-group factor. That significance can be attributed to lower performance under the slower frequencies (1 & 2 Hertz) and the smaller amplitudes (25% & 50%) for the small to large text size comparisons. Blink amplitudes need to be 75% amplitude or greater to be maximally attention-getting if the text is small (near .10 inch). Blink amplitudes of 50% or more were sufficient with large text (near .20 inch).

DISCUSSION

The 75% and 100% amplitude conditions consistently produced the lowest errors regardless of the blink frequencies and text sizes used in this experiment. Furthermore, in Step 4 of the model comparison procedure, involving the 2, 3, and 4 HZ frequencies, 3 text sizes, and only the 75% and 100% amplitudes; the small performance differences among

the conditions for both dependent variables are probably within tolerable real-world limits for most practical applications. As discussed above, mean differences of less than .7 sec are probably meaningless in most gross motor tasks. Cell means for both dependent measures are plotted for this combination of parameters on Figure 7. The figure indicates stable performance under these parameters.

CONCLUSIONS

Our results support the use of amplitudes 75% or greater with frequencies from 2 to 4 Hertz and text size 0.15 inch or greater for optimum highlighting value in visual search tasks.

Our conservative recommendations for optimal blink characteristics (among those tested in this experiment) were based on the following criteria:

1) Maximum accuracy, minimum response times, and standard deviations of participants observed under each experimental condition
2) Knowledge of minimum response times with optimal target highlighting (e.g., under the control condition)
3) Statistical significance of performance differences
4) Practical implications of observed differences with regard to real-world applications
5) Preference of erring on the side of being too conservative and cautious especially in safety-critical applications

The reader is encouraged to make blink parameter selection decisions appropriate for particular applications.

To summarize, research has demonstrated that blinking is beneficial in reducing search time in monitoring tasks. Furthermore, blinking has an added benefit for attracting attention given particular parametric values of blink rate, amplitude, duration, and target size. Knowledge of those parameters may be useful whether blink is used as a single design feature or as a redundant cue. For example, when color is designed to facilitate task performance by attracting attention to information, and blinking is used as a redundant cue for color deficients. Blinking can be used in many forms on modern CRT displays and can have many real-world applications. Traditionally, blinking has been used to convey warning — such as traffic lights, or urgency — such as on fire trucks and police cars. To some degree, that precept or assumption translates across cultures. So, even though this study examines blink in a somewhat limited context, broader application of these findings may be pertinent.

REFERENCES

Anastasi, D.L., Hill, B.P., Murphy, E.D., Cardosi, K.M., Guttman, J.A., & Amaldi, P. (1995). *Computer-Human Interface (CHI) Considerations in Human Factors in the Design and Evaluation of Air Traffic Control Systems.* DOT/FAA/RD-95/3. Office of Aviation Research, Washington, DC 20591.

Bailey, R.W. (1989). Human performance engineering: using human factors/ergonomics to achieve computer system usability. (2nd Ed.) New Jersey. Prentice-Hall.

Boff, K.R. & Lincoln, J.E. (Eds.). (1988). *Engineering Data Compendium: Human Perception and Performance.* AAMRL, Wright-Patterson AFB, OH

Christ, R.E. (1975). Review and analysis of color coding research for visual displays. *Human Factors.* 17(6) 542-70.

Gerathewohl, S.J. (1951). Conspicuity of flashing and steady light signals. I. Variation of contrast. USAF School of Aviation Medicine, Special Report.

Gerathewohl, S.J. (1952). Conspicuity of flashing and steady light signals. II. High contrasts. USAF School of Aviation Medicine, Special Report.

Gerathewohl, S.J. (1953). Conspicuity of flashing and steady light signals: Variation of contrasts. *Journal of Optical Society of America.* 43, 567-71.

Gerathewohl, S.J. (1954). Conspicuity of flashing light signals of different frequency and duration. *Journal of Experimental Psychology.* 48 (4) 247-51.

Gilmore, W.E., Gertman, D.I., and Balckman, H.S. (1989). *The user-computer interface in process control: A human factors engineering handbook.* New York: Academic Press.

Mertens, H.W. & Milburn, N.J. (1998). Effects of color vision deficiency on detection of highlighted targets in a simulated air traffic control display. Unpublished manuscript.

Milburn, N.J. & Mertens, H.W. (1997). Evaluation of a range of target blink amplitudes for attention-getting value in a simulated air traffic control display. Federal Aviation Administration (FAA) Office of Aviation Medicine. Report No. DOT/FAA/AM-97/10. Washington, DC. U.S. Department of Transportation. NTIS #ADA326465.

Military Standards, MIL-STD-1472D, (1989). Human Engineering Design Criteria for Military Systems, Equipment and Facilities.

Society of Automotive Engineers, SAE G-10 Committee, Silverstein, L.D., Jacobesen, A., Post, D., Druk, R.V., Mertens, H.W., & Gershzohn, G. (1988). Human engineering considerations in the application of color to electronic aircraft displays. SAE Technical Report ARP4032. Warrendale, PA: Society of Automotive Engineers, Inc.

Silverstein, L.D. & Merrifield, R.M. (1985). The development and evaluation of color systems for airborne applications. Phase 1: Fundamental visual, perceptual, and display system considerations. Technical Report DOT/FAA/PM-85-19. Washington, DC: U.S. Department of Transportation.

Thackray, R.I. & Touchstone, R.M. (1991). Effects of monitoring under high and low taskload on detection of flashing and coloured radar targets. Special Issue: Cognitive ergonomics II. *Ergonomics. 34(8)* 1065-81.

Van Orden, K.F. & Di Vita, J. (1993). Highlighting with flicker: Some basic visual search considerations. *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting—1993.* 1300-4.

APPENDIX A

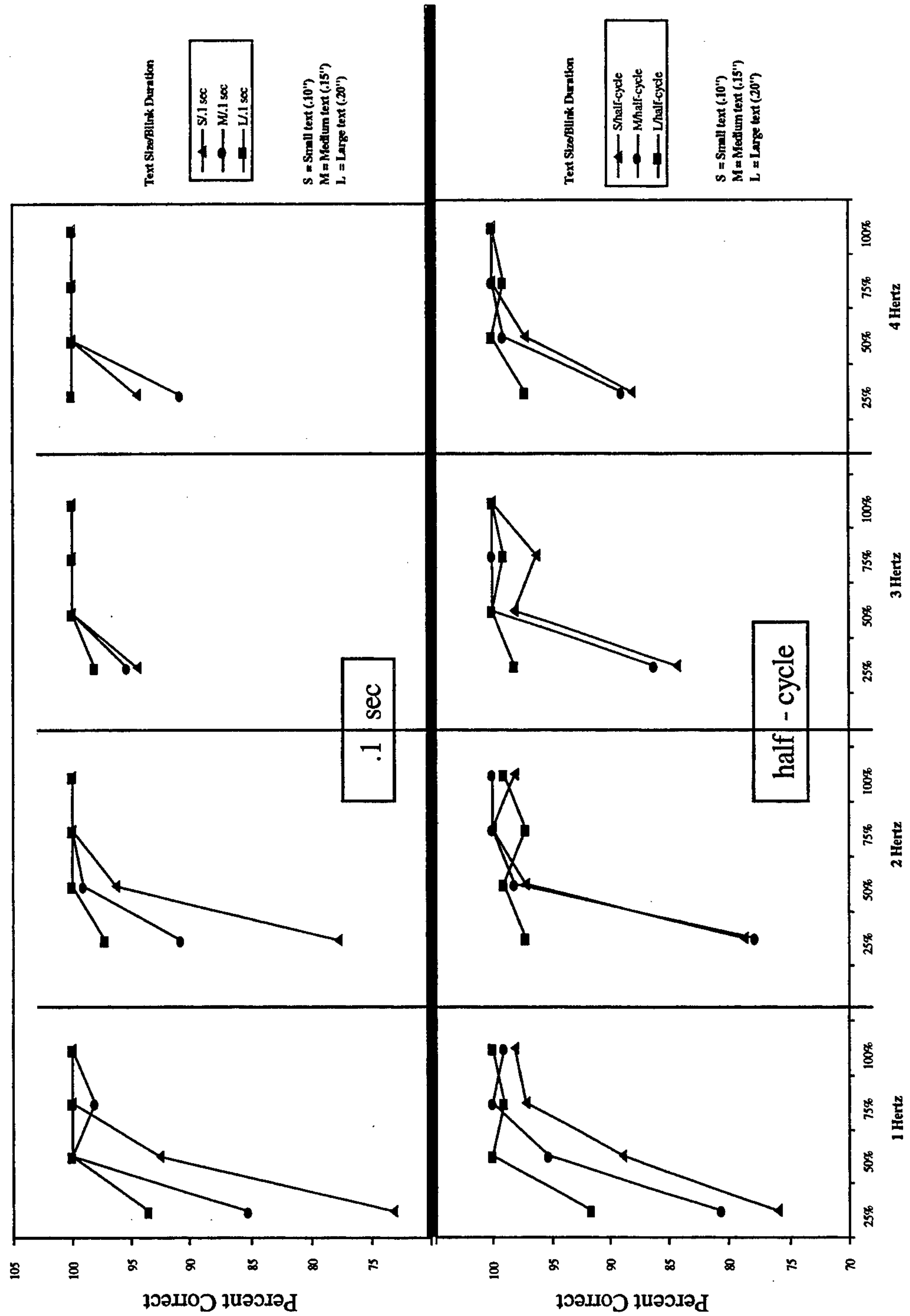

Figure 1. Percent correct as a function of frequency, amplitude, text size, and blink duration

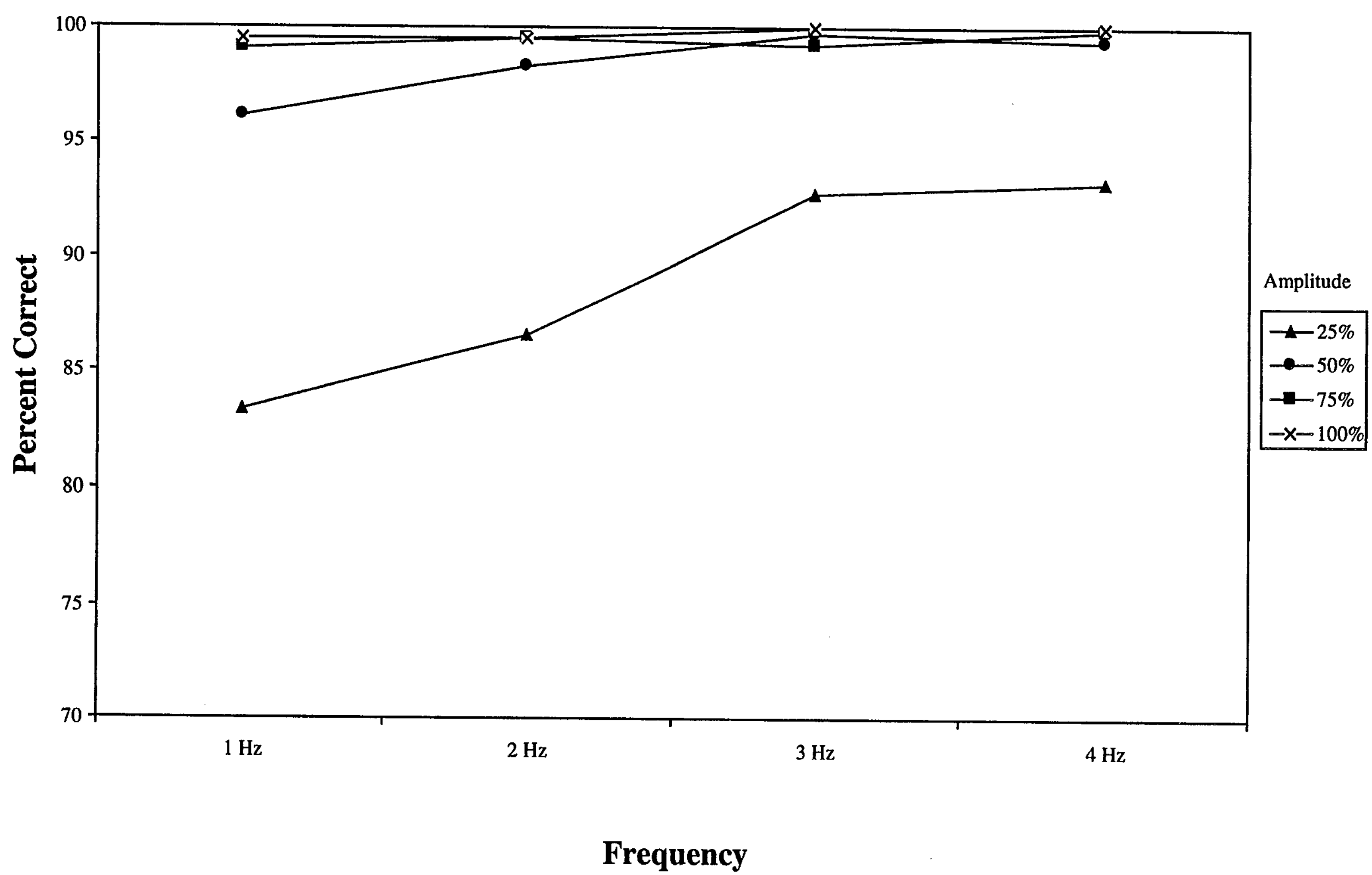

Figure 2. Percent correct as a function of frequency and amplitude

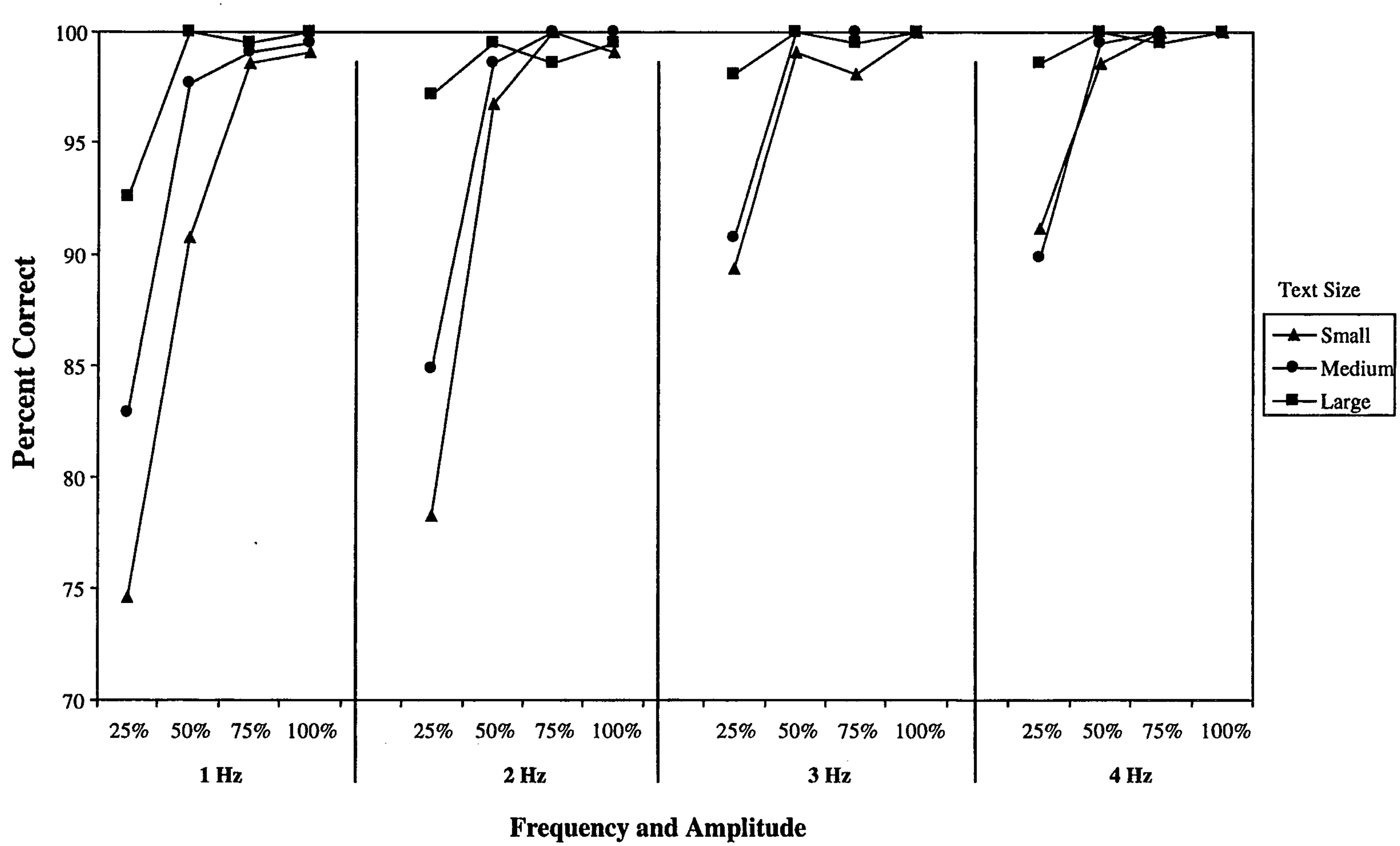

Figure 3. Percent correct as a function of frequency, amplitude, and text size

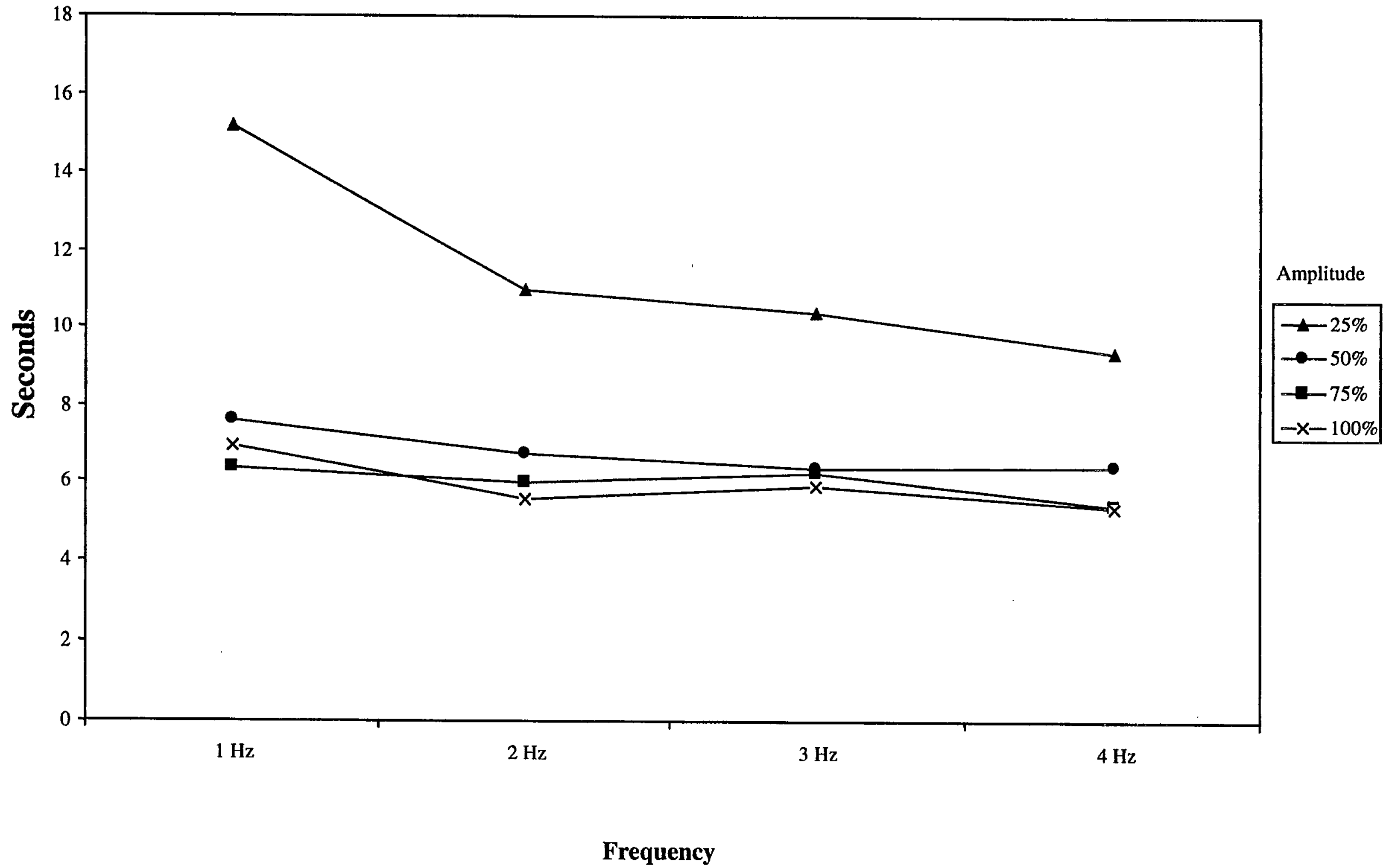

Figure 4. Response time as a function of frequency and amplitude

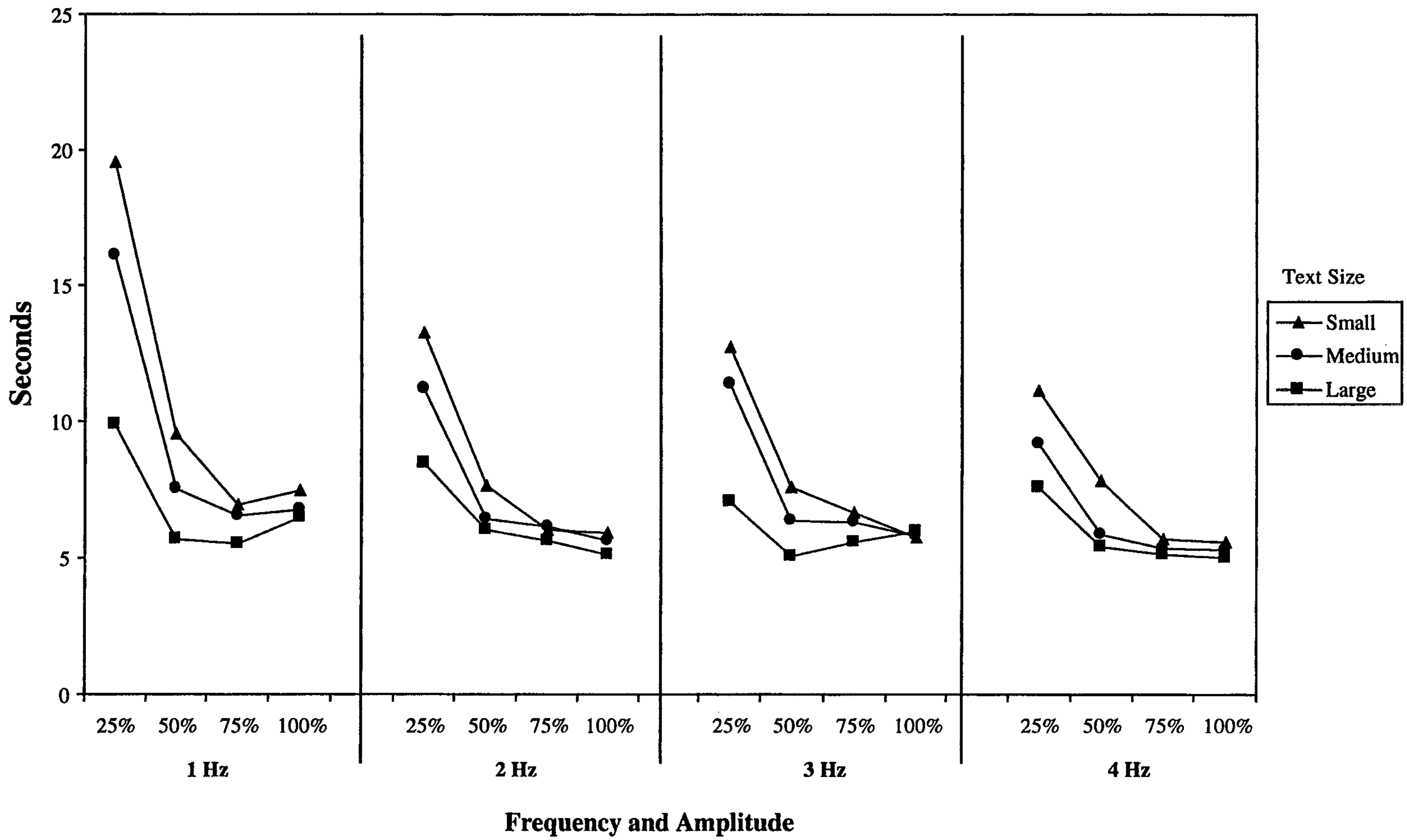

Figure 5. Response time as a function of frequency, amplitude, and text size

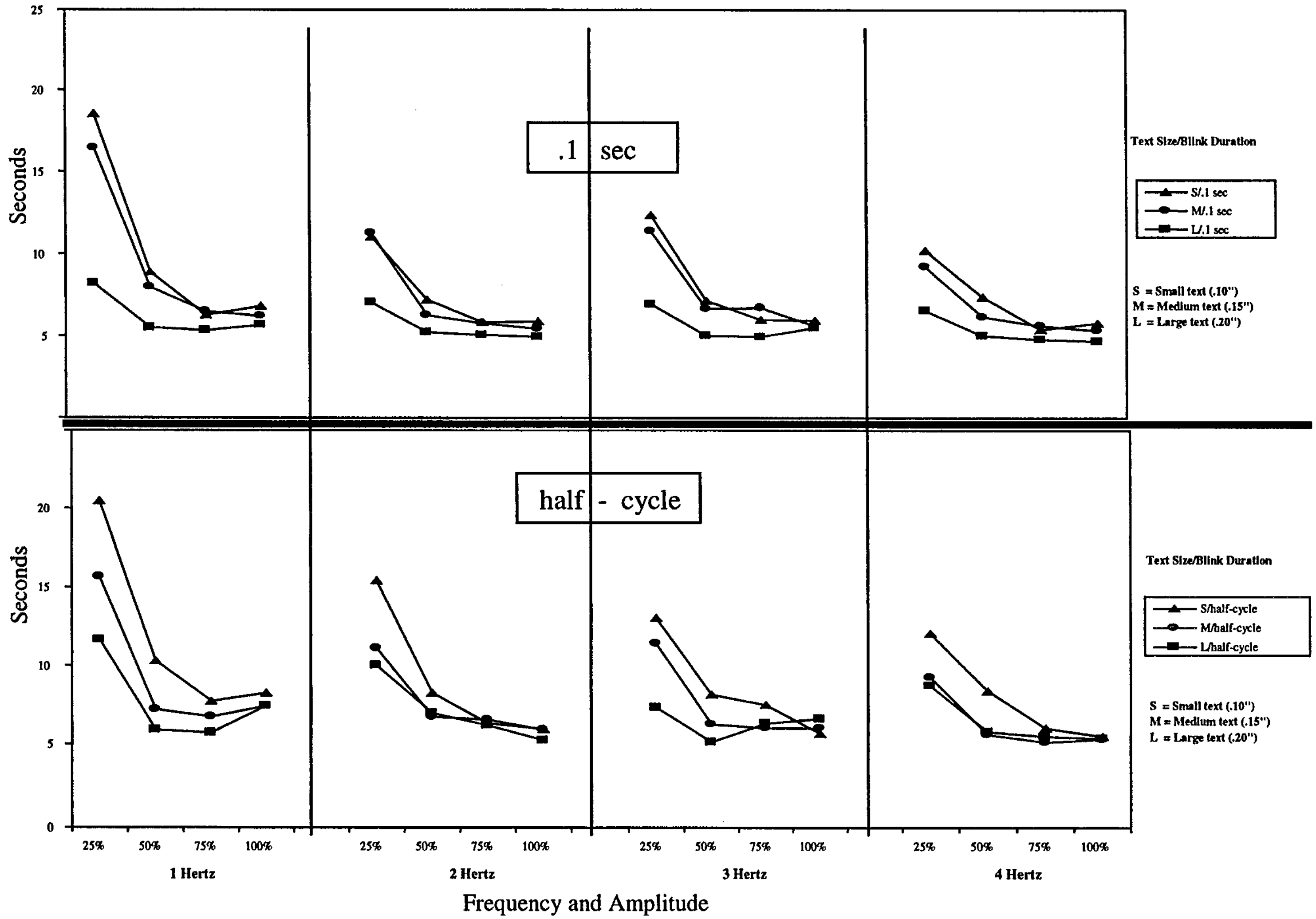

Figure 6. Response time as a function of frequency, amplitude, text size, and blink duration

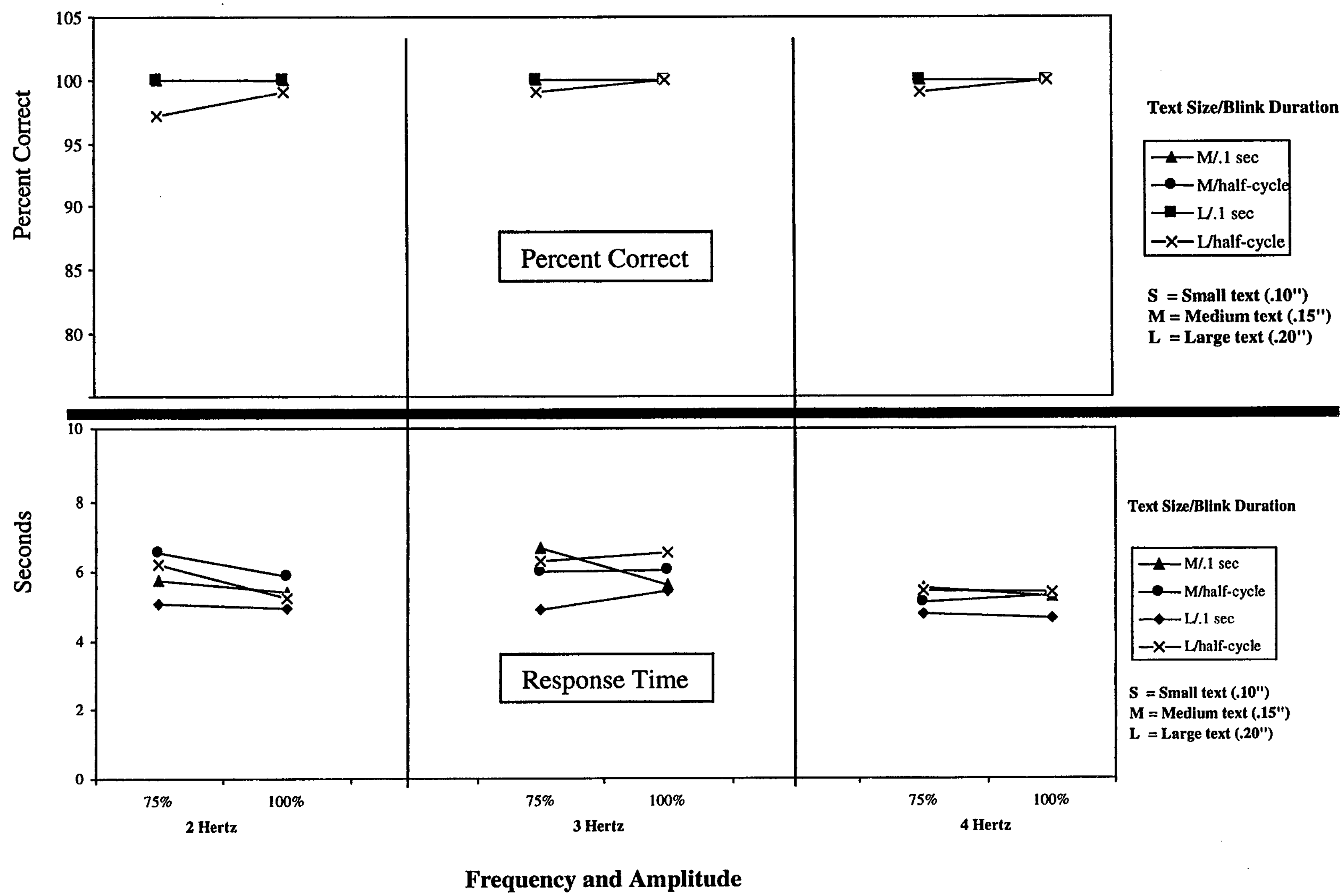

Figure 7. Response time and percent correct as a function of frequency, amplitude, text size, and blink duration

APPENDIX B

Table 1: Age of Participants with Cell and Marginal Means

Half-Cycle	**.1Sec**	**Marginal Size Means**
	Small .10 inch	
18	19	
20	19	
20	21	
25	23	
29	25	
31	30	
$\bar{X} = 23.83$	$\bar{X} = 22.83$	$\bar{X} = 23.33$
	Medium .15 inch	
19	18	
19	19	
21	19	
21	25	
27	28	
34	30	
$\bar{X} = 23.50$	$\bar{X} = 23.17$	$\bar{X} = 23.33$
	Large .20 inch	
18	18	
21	22	
22	23	
31	23	
34	34	
34	34	
$\bar{X} = 25.16$	$\bar{X} = 25.83$	$\bar{X} = 25.50$
Half-Cycle mean	**.1sec mean**	**Grand Mean**
$\bar{X} = 24.17$	$\bar{X} = 23.94$	$\bar{X} = 24.0556$

Table 2: Design of Study

	DURATION	.1 sec (n = 18)															
	FREQUENCY	1 HERTZ				2 HERTZ				3 HERTZ				4 HERTZ			
	AMPLITUDE (IN PERCENT)	25	50	75	100	25	50	75	100	25	50	75	100	25	50	75	100
TEXT SIZE	SMALL n = 6																
	MEDIUM n = 6																
	LARGE n = 6																

	DURATION	Half-Cycle (n = 18)															
	FREQUENCY	1 HERTZ				2 HERTZ				3 HERTZ				4 HERTZ			
	AMPLITUDE (IN PERCENT)	25	50	75	100	25	50	75	100	25	50	75	100	25	50	75	100
TEXT SIZE	SMALL n = 6																
	MEDIUM n = 6																
	LARGE n = 6																

Table 3: Control Condition Mean Response Times for Subjects in Each of the Between-Subject Groups

Size		Duration		
		.1 SEC	HALF-CYCLE	MARGINAL MEANS
SMALL	.10"	4.96	5.23	5.10
MEDIUM	.15"	5.10	4.72	4.91
LARGE	.20"	3.98	4.50	4.24
MARGINAL MEANS		4.68	4.82	4.75

*This table is provided to show equivalent performance for subjects in each of the 6 groups for between-subjects analysis in the control condition.

Table 4: Descriptive Statistics for Total Percent Correct as a Function of Text Size

		Small			*Medium*			*Large*		
AMPLITUDE	HZ	MIN	MEAN	SD	MIN	MEAN	SD	MIN	MEAN	SD
25% Amp	1	33.33	74.53	23.14	50.0	82.87	17.32	72.2	92.59	9.27
	2	22.22	78.24	30.00	22.22	84.87	22.07	83.33	97.22	5.55
	3	38.89	89.35	16.99	27.78	90.74	20.42	88.89	98.14	4.32
	4	27.78	91.20	20.44	55.56	89.81	14.57	88.89	98.61	3.45
50% Amp	1	61.11	90.74	11.19	83.33	97.68	5.00	100.0	100.0	0
	2	83.33	96.75	5.53	94.44	98.61	2.51	94.44	99.53	1.6
	3	88.89	99.07	3.20	100.0	100.0	0	100.0	100.0	0
	4	88.89	98.61	3.45	94.44	99.53	1.60	100.0	100.0	0
75% Amp	1	88.89	98.61	3.45	94.44	99.07	2.16	94.44	99.53	1.6
	2	100.0	100.0	0	100.0	100.0	0	83.33	98.61	4.81
	3	83.33	98.14	4.93	100.0	100.0	0	94.44	99.53	1.60
	4	100.0	100.0	0	100.0	100.0	0	94.44	99.53	1.60
100% Amp	1	94.44	99.07	2.16	94.44	99.53	1.60	100.0	100.0	0
	2	94.44	99.07	2.16	100.0	100.0	0	94.44	99.53	1.60
	3	100.0	100.0	0	100.0	100.0	0	100.0	100.0	0
	4	100.0	100.0	0	100.0	100.0	0	100.0	100.0	0

* On all conditions at least one person achieved a maximum of 100% correct; therefore, the maximum range is omitted from this table.

* Cells in which the minimum performance is less than 90% correct have been shaded on Table 4 to serve as visual aid relevant to the task of establishing a cut-off point for purposes of making recommendations for blink parameters.

* N = 12 in each text size group

Table 5: Percent Correct as a Function of Duration Under the 2, 3, and 4 Hertz and Amplitudes of 75 and 100%

DURATION	FREQUENCY/AMPLITUDE					
	2 HERTZ		3 HERTZ		4 HERTZ	
	75%	100%	75%	100%	75%	100%
.1 SEC	100	100	100	100	100	100
HALF-CYCLE	95	83	72	100	94.5	100

* [$F_{1,30}$=4.31, p=.047]

Table 6: Model Comparisons and F-Statistics for Accuracy Analysis

Step 1				*Step 2*				*Step 3*				*Step 4*			
FREQ	AMP	SIZE	DUR	FREQ	AMP	SIZE	DUR	FREQ	AMP	SIZE	DUR	FREQ	AMP	SIZE	DUR
1 Hz	25%	.10	.1 sec	1 Hz	25%	.10	.1 sec	1 Hz	25%	.10	.1 sec	1 Hz	25%	.10	.1 sec
2 Hz	50%	.15	half-cycle	2 Hz	50%	.15	half-cycle	2 Hz	50%	.15	half-cycle	2 Hz	50%	.15	half-cycle
3 Hz	75%	.20		3 Hz	75%	.20		3 Hz	75%	.20		3 Hz	75%	.20	
4 Hz	100%			4 Hz	100%			4 Hz	100%			4 Hz	100%		

Step 1	*Step 2*	*Step 3*	*Step 4*
MULTIVARIATE TESTS	MULTIVARIATE TESTS	MULTIVARIATE TESTS	MULTIVARIATE TESTS
Hz [$F_{3,28}$=7.65, p = .001] Amp [$F_{3,28}$= 8.16, p < .001] Hz*Amp [$F_{9,22}$ = 3.04, p = .016]	Hz [$F_{3,28}$= 5.05, p = .006] Amp [$F_{1,30}$ = 4.10, p = .052] Hz*Amp [$F_{3,28}$ = 4.64, p = .009]	Hz*Amp [$F_{2,29}$ = 4.20, p = .025]	None
BETWEEN	BETWEEN	BETWEEN	BETWEEN
Size [$F_{2,30}$= 3.50, p = .043]	Size [$F_{2,30}$= 4.89, p = .015]	None	Duration [$F_{1,30}$= 4.31, p = .047]
MULTIPLE COMPARISONS	MULTIPLE COMPARISONS	MULTIPLE COMPARISONS	MULTIPLE COMPARISONS
25% Amp, 1 Hz small vs. large text size 50% Amp, 1 Hz small vs. large text size	50% Amp, 1Hz small vs. large text size	None	None

Table 7: Descriptive Statistics for Total Response Time as a Function of Text Size

		Small				*Medium*				*Large*			
AMP.	**Hz**	**MAX**	**MIN**	**MEAN**	**SD**	**MAX**	**MIN**	**MEAN**	**SD**	**MAX**	**MIN**	**MEAN**	**SD**
Control	N/A	8.42	3.18	5.10	1.76	6.29	3.79	4.91	.78	7.36	3.22	4.24	1.16
25% Amp	1	31.61	7.26	19.54	7.64	33.79	6.09	16.06	7.93	19.02	4.71	9.91	4.17
	2	19.43	5.62	13.24	4.54	21.50	6.37	11.17	4.36	19.98	4.24	8.51	4.26
	3	18.95	7.57	12.70	3.76	25.64	5.24	11.35	6.54	13.31	3.73	7.10	2.67
	4	16.75	5.52	11.11	3.44	15.22	5.01	9.22	3.31	17.91	3.97	7.62	4.11
50% Amp	1	14.39	5.09	9.57	2.77	10.40	5.63	7.56	1.60	7.58	4.53	5.70	1.05
	2	10.87	4.30	7.71	2.11	8.56	4.59	6.48	1.44	16.82	4.02	6.06	3.56
	3	10.75	4.11	7.62	2.03	9.07	4.79	6.43	1.42	7.86	3.76	5.06	1.27
	4	15.14	4.92	7.84	3.01	7.95	4.37	5.89	1.04	11.08	3.65	5.40	2.19
75% Amp	1	12.08	4.67	6.98	1.94	8.03	4.95	6.59	1.01	6.93	4.26	5.53	.96
	2	9.55	4.32	6.06	1.61	7.67	4.53	6.15	1.21	12.41	3.76	5.64	2.34
	3	15.39	4.10	6.71	2.89	9.34	4.77	6.33	1.54	14.91	3.59	5.62	3.07
	4	10.20	4.02	5.69	1.67	7.76	4.21	5.35	1.03	10.35	3.42	5.12	2.00
100% Amp	1	16.19	4.94	7.52	3.21	11.67	5.30	6.80	1.76	17.53	4.38	6.55	3.63
	2	9.38	4.30	5.94	1.68	8.10	4.61	5.66	.98	7.00	3.81	5.10	1.07
	3	10.74	3.98	5.77	1.98	8.08	3.69	5.83	1.41	15.07	3.68	6.02	3.44
	4	8.22	4.26	5.62	1.32	6.64	4.05	5.32	.89	8.84	3.66	5.03	1.51

* Cells in which the mean response time is greater than the 95th percentile (8.16 sec) for the control condition have been shaded on Table 7 to provide a visual aid relevant to the task of establishing a cut-off point for purposes of making recommendations for blink parameters.

* N = 12 in each size group

Table 8: Model Comparisons and F-Statistics for Response Time

Step 1				Step 2				Step 3				Step 4			
FREQ	AMP	SIZE	DUR	FREQ	AMP	SIZE	DUR	FREQ	AMP	SIZE	DUR	FREQ	AMP	SIZE	DUR
1 Hz	25%	.10	.1 sec	1 Hz	25%	.10	.1 sec	1 Hz	25%	.10	.1 sec	1 Hz	25%	.10	.1 sec
2 Hz	50%	.15	half-cycle	2 Hz	50%	.15	half-cycle	2 Hz	50%	.15	half-cycle	2 Hz	50%	.15	half-cycle
3 Hz	75%	.20		3 Hz	75%	.20		3 Hz	75%	.20		3 Hz	75%	.20	
4 Hz	100%			4 Hz	100%			4 Hz	100%			4 Hz	100%		

	Step 1	Step 2	Step 3	Step 4
Multivariate Tests	Hz [$F_{3,28}$ = 49.23, p < .001] Hz*Size [$F_{6,58}$ = 3.51, p = .005] Amp [$F_{3,28}$ = 22.79, p < .001] Amp*Size [$F_{6,58}$ = 5.02, p < .001] Hz*Amp [$F_{9,22}$ = 4.65, p < .002]	Hz [$F_{3,28}$ = 10.90, p < .001] Amp [$F_{1,30}$ = 29.09, p < .001] Amp*Size [$F_{2,30}$ = 12.84, p < .001] Hz*Amp [$F_{3,28}$ = 3.45, p = .03]	Hz [$F_{2,29}$ = 5.75, p = .008] Amp [$F_{1,30}$ = 20.92, p < .001] Amp*Size [$F_{2,30}$ = 11.062, p < .001]	Hz [$F_{2,29}$ = 12.36, p < .001]
Between	Size [$F_{2,30}$ = 3.66, p = .038]	None	None	None
Multiple Comparisons	25% Amp, 1 Hz small vs. large 50% Amp, 1 Hz small vs. large 50% Amp, 1 Hz med. vs. large 75% Amp, 1 Hz med. vs. large 25% Amp, 2 Hz small vs. large 25% Amp, 3 Hz small vs. large 50% Amp, 3 Hz small vs. large	50% Amp, 1 Hz small vs. large 50% Amp, 1 Hz med. vs. large 75% Amp, 1 Hz med. vs. large 50% Amp, 3 Hz small vs. large 50% Amp, 3 Hz med. vs. large	50% Amp, 1 Hz small vs. large 50% Amp, 1 Hz med. vs. large	None